DOCUMENT RESUME

ED 236 212 TM 830 725

AUTHOR Argulewicz, Ed N.; Abel, R. Robert
TITLE Item Bias in the PPVT-R for Anglo-American and Mexican-American Children.
PUB DATE Aug 83
NOTE 29p.; Paper presented at the American Psychological Association Annual Convention (Anaheim, CA, August 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Variance; Content Analysis; *Diagnostic Tests; Elementary Education; *Hispanic Americans; Research Methodology; *Test Bias; Testing Problems; *Test Items; *White Students
IDENTIFIERS *Peabody Picture Vocabulary Test (Revised)

ABSTRACT

This study examines item bias on Forms L and M of the Peabody Picture Vocabulary Test-Revised (PPVT-R) for 132 Anglo-American and Mexican-American children in grades one through four. Analyses of variance were employed to assess item bias as defined by items x ethnicity interactions. Follow-up analyses were performed using a Bonferroni-type procedure on individual item contrasts. Bias as measured by differences in item difficulty was found for both groups; however, there was no clear pattern to the items that were more difficult for either group. The small number of items that were more difficult for one ethnic group than the other, coupled with the high reliability of performance overall for both groups, suggest that bias in content of the PPVT-R is minimal. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236212

Item Bias in the PPVT-R for Anglo-American and Mexican-American Children

Ed N. Argulewicz and R. Robert Abel

Arizona State University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E.N. Argulewicz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: PPVT-R Item Bias

Paper presented at the American Psychological Association Annual Convention, Anaheim, California, August 1983.

TM 830 225

Abstract

This study examines item bias on Forms L and M of the Peabody Picture Vocabulary Test-Revised (PPVT-R) for a sample of Anglo-American and Mexican-American children. Analyses of variance were employed to assess item bias as defined by items x ethnicity interactions. Follow-up analyses were performed using a Bonferroni-type procedure on individual item contrasts. Bias as measured by differences in item difficulty was found for both groups; however, there was no clear pattern to the items that were more difficult for either group. The small number of items that were more difficult for one ethnic group than the other, coupled with the high reliability of performance overall for both groups suggest that bias in content of the PPVT-R is minimal.

Item Bias in the PPVT-R for Anglo-American and Mexican-American Children

The Peabody Picture Vocabulary Test (PPVT) (Dunn, 1959) has been widely used by psychologists, speech pathologists and educators. The test's uses include screening retardates, preschoolers, and gifted children, and evaluating treatment programs as a pre-posttest measure. The PPVT was recently revised (PPVT-R) (Dunn & Dunn, 1981), and the current version shows marked improvements over its predecessor including a more representative standardization sample, elimination of items which stereotype sex roles, and greater representation of minorities in the test plates.

With these modifications, questions arise as to changes in the psychometric characteristics of the test. To date, there have been a number of studies reported regarding the PPVT-R's validity (Argulewicz, Bingenheimer, & Anderson, in press; Bracken & Prasse, 1982; Bracken & Prasse, in press; Breen, 1981; McCallum, Karnes, & Bracken, in press; McCallum, Mimms, & Bracken, in press; Naglieri, 1981; Prasse & Bracken, 1981), and reliability (Argulewicz, Anderson, & Bingenheimer, in press; Bracken & Prasse, 1981; McCallum & Bracken, 1981; Stoner, 1981). In all of these studies, external criteria have always been used for determining the validity of the PPVT-R and alternate form equivalence has been the method for ascertaining reliability estimates. There are currently no empirical data which examine the internal performance of groups of the PPVT-R. Such data are a particularly useful source for evaluating the existence of bias in tests (Jensen, 1980).

The present study examines the internal performance of Anglo-American and Mexican-American children. The information obtained can be used in evaluating item bias on Forms L and M of the PPVT-R.

Subjects and Procedures

The sample consisted of 132 students in grades one through four in a low to mid-SES urban school district in the southwest. Students were selected on a stratified random basis in order to approximate the percentages of both ethnic groups as they existed in the total district population. No attempt was made to eliminate students from the selected sample who were enrolled in special education resource programs. Of the total, 91 were Anglo Americans (47 males and 44 females) and 41 were Mexican Americans (18 males and 23 females). Mean ages were 106 months (SD = 15) for the Anglo students and 103 months (SD = 14) for the Mexican-American students. All of the Mexican-American students came from homes in which Spanish was the primary language.

All students were individually administered both forms of the PPVT-R in a counterbalanced order. Thus, half of the students were administered Form L followed by Form M, and half were administered Form M followed by Form L. All testing took place in the spring of the school year.

Data Analysis

Data from Forms L and M were analyzed separately. For both test forms, items 1 through 19 and 141 through 175 were excluded from analysis due to lack of variation. All children passed items 1 through 19 and all children failed items 141 through 175 on both forms. Separate 2 (ethnicity: Anglo- vs. Mexican-American) x 4 (grades 1 through 4) x 120 (PPVT-R

test items 20 through 140) repeated measures ANOVAs were conducted on arcsin transformed item difficulties from each test form. In these analyses, we were interested only in the items x ethnicity interaction. When a significant items x ethnicity interaction was found, we performed interaction contrasts on the individual item difficulties (Plake & Hoover, 1979). These contrasts identified items which produced differential performance between the ethnic groups. Because of the large number of contrasts necessary, we used a Bonferroni-type procedure to maintain an experiment-wise Type I error rate at $\alpha = .05$. Contrasts whose value exceeded the Bonferroni critical value were taken to indicate bias on their respective item.

Results

Means, standard deviations, and reliability estimates by ethnicity, grade, and gender for Forms L and M of the PPVT-R are presented in Table 1. Although not a primary focus of this study, differences in group performance on the two forms were compared by an ethnicity x grade x gender MANOVA. Results of this analysis indicated a significant overall effect for ethnicity $F(2, 115) = 8.78$, $p < .0004$, but grade and gender effects were nonsignificant. The only significant interaction effect was found for the ethnicity x grade x sex term $F(6, 230) = 2.38$, $p < .03$.

Follow-up univariate ANOVAs showed that Anglos scored significantly higher than Mexican Americans on Form L $F(1, 116) = 17.20$, $p < .0001$, and Form M $F(1, 116) = 10.11$, $p < .002$, and that there was a significant ethnicity x grade x sex interaction effect for only Form L $F(3, 116) = 3.33$, $p < .03$. Simple effects tests indicated that this interaction was

the result of a sex difference within first-grade Anglo students in which females ($\bar{X}$ = 108.14, SD = 14.88) scored significantly higher than males ($\bar{X}$ = 95.00, SD = 13.25) $F(1, 116 = 3.42, p < .05)$.

Insert Table 1 about here

The significance of the difference between the reliabilities for each ethnicity, grade, and gender grouping was tested via a technique developed by Feldt (1969) and adapted by Reynolds (1982). This technique involves dividing the largest error variance by the smallest error variance to compute an F max statistic and comparing the obtained F to a Bonferroni adjusted critical value. Using this technique, no significant differences were found among any of the internal consistency estimates.

Items x ethnicity statistics are presented in Table 2. Repeated measures analysis of variance revealed significant items x ethnicity interactions for both forms. However, it should be noted that the ANOVA methodology overestimates the degree of bias present when two groups differ on total test score.

Insert Table 2 about here

In order to identify the items contributing to these interactions, we performed 120 contrasts on the Form L and Form M arcsin transformed item difficulties. These contrasts illustrated that proportionately more Anglo- than Mexican-American students passed items 56, 62, 70, 71, 72, 84,

and 86 on Form L and items 70, 77, 79, 81, and 88 on Form M. Significantly more Mexican Americans than Anglos passed items 59 and 86 on Form L and item 61 on Form M.

Although the items x ethnicity interactions were significant beyond the .001 probability level for both PPVT-R forms, the amount of variance accounted for of the total variation by these effects was nominal. In Form L, the items x ethnicity interaction accounted for .0099 of the total variance, and .0075 of the total variance in Form M. It is doubtful that any source of variance accounting for less than 1% of a test's total variation could disrupt the test's properties of internal consistency.

Discussion

The findings of significant items x ethnicity interactions and high reliability estimates for both Anglo-American and Mexican-American group performance on the PPVT-R are consistent with Jensen's (1974) study conducted with the PPVT-R's predecessor, the PPVT, as well as studies on other verbal aptitude measures (Reynolds, 1982). Although the items x ethnicity interactions were highly significant, the minuscule percentage of total variance accounted for in all likelihood indicates little practical significance.

Discerning the nature of the biased items is, at present, a speculative task. One potential source of item bias may be the location of the item within the test (Dwyer, Note 1). Examination of the biased items reveals no apparent trend in regard to item location. Items that favor Anglo-American children on one form are either adjacent to, or in very

close proximity to, items that favor the Mexican-American children on the alternate form.

Examination of the individual items identified as biased is no more productive. When parts of speech are analyzed, there are eleven nouns and three verbs in which bias was identified. Eight of the nouns were easier for the Anglo-American group and three were easier for the Mexican-American group. All three of the verbs were easier for the Anglo-American group. Unfortunately, this small number of items of both word functions, as well as lack of an identifiable consistent pattern of performance, precludes meaningful interpretation. Additionally, the small sample size limits the reliability of these results.

The absence of a unidirectional trend in item bias in a statistical sense, along with a lack of any identifiable pattern in bias from a post-hoc item inspection, suggests that there does not exist a systematic bias in PPVT-R item content. However, the question of how the elimination of specific items in which there was differential performance between the two ethnic groups influences concurrent and predictive validity warrants further investigation.

Reference Note

1. Dwyer, C. A. Test content in mathematics and science: The consideration of sex. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

References

Argulewicz, E. N., Anderson, C. C., & Bingenheimer, L. T. Alternate form reliability of the PPVT-R for Anglo American and Mexican American school children. Educational and Psychological Research, in press.

Argulewicz, E. N., Bingenheimer, L. T., & Anderson, C. C. Comparison of concurrent validity of the PPVT-R for Anglo American and Mexican American students. Journal of Psychoeducational Assessment, in press.

Bracken, B. A., & Prasse, D. P. Alternate form reliability of the PPVT-R for white and black educable mentally retarded students. Educational and Psychological Research, 1981, 1, 151-154.

Bracken, B. A., & Prasse, D. P. Comparison of the PPVT, PPVT-R, and intelligence tests used for placement of black, white, and Hispanic EMR students. Journal of School Psychology, 1982, 19, 304-311.

Bracken, B. A., & Prasse, E. P. Concurrent validity of the PPVT-R for "at-risk" preschool children. Psychology in the Schools, 1983, 20, 13-15.

Breen, M. J. Comparison of the Wechsler Intelligence Scale for Children-Revised and the Peabody Picture Vocabulary Test-Revised for a referred population. Psychological Reports, 1981, 49, 717-718.

Dunn, L. M. Peabody Picture Vocabulary Test. Minneapolis, Minn.: American Guidance Service, 1959.

Dunn, L. M., & Dunn, L. M. Peabody Picture Vocabulary Test-Revised. Circle Pines, Minn.: American Guidance Service, 1981.

Feldt, L. S. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. Psychometrika, 1969, 34, 363-373.

Jensen, A. R. How biased are culture loaded tests? Genetic Psychology Monographs, 1974, 90, 185-244.

Jensen, A. R. Bias in mental testing. New York: Free Press, 1980.

McCallum, R. S., & Bracken, B. A. Alternate form reliability of the PPVT-R for white and black preschool children. Psychology in the Schools, 1981, 18, 422-425.

McCallum, R. S., Karnes, F., & Bracken, B. A. Comparison of the PPVT and PPVT-R, and the WISC-R with gifted students. Journal for Education of the Gifted, in press.

McCallum, R. S., Mimms, M., & Bracken, B. A. Comparability of the Peabody Picture Vocabulary Test, Peabody Picture Vocabulary Test-Revised, and Stanford-Binet for trainable mentally retarded persons. Educational and Psychological Research, 1982, 2, 255-260.

Naglieri, J. A. Concurrent validity of the Revised Peabody Picture Vocabulary Test. Psychology in the Schools, 1981, 19, 286-289.

Plake, B. S., & Hoover, H. D. An analytic method for identifying biased items. Journal of Experimental Education, 1979, 48, 153-154.

Prasse, D. P., & Bracken, B. A. Comparison of the PPVT-R and WISC-R with urban educable mentally retarded students. Psychology in the Schools, 1981, 18, 174-177.

Reynolds, C. R. Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press, 1982.

Reynolds, C. R. The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology. New York: Wiley, 1982.

Stoner, S. B. Alternate form reliability of the Revised Peabody Picture Vocabulary Test for head start children. Psychological Reports, 1981, 49, 628.

Winer, B. T. Statistical procedures in experimental design. New York: McGraw-Hill, 1971.

Table 1

Means, Standard Deviations, and KR-20 Reliabilities on PPVT-R Forms L and M by Ethnicity, Grade, and Gender

	Form L			Form M		
	$\bar{X}$	SD	KR-20	$\bar{X}$	SD	KR-20
Ethnicity						
Anglo Americans (N = 91)	98.40	14.39	.96	94.57	16.13	.95
Mexican Americans (N = 41)	87.10	14.70	.95	86.10	11.71	.94
Grade						
First (N = 21)	98.05	17.04	.95	99.48	15.40	.95
Second (N = 21)	93.00	14.05	.89	87.70	21.53	.91
Third (N = 47	93.23	16.80	.94	91.08	12.28	.91
Fourth (N = 43)	96.07	13.44	.93	91.19	14.34	.94
Gender						
Males (N = 65)	96.45	14.71	.96	92.52	17.78	.94
Females (N = 67)	93.37	15.92	.95	91.37	12.72	.95

Table 2

Repeated Measures Analysis of Variance

on PPVT-R Forms L and M

Form	Source of Variance	df	SS	MS	F
L	Items	119	1416.09	11.90	153.79**
	Items x Grade	357	72.89	.20	2.64**
	Items x Ethnicity	119	15.89	.13	1.73**
	Error	357	27.62	.08	
M	Items	119	1573.87	13.23	194.00**
	Items x Grade	357	54.94	.15	2.26**
	Items x Ethnicity	119	12.82	.11	1.58*
	Error	357	24.28	.07	

Note. Since each cell in this design contained only a single observation, the within-cells mean square was unsuitable as an error term. Instead, the items x grade x ethnicity interaction was used (Winer, 1971).

*$p < .001$

**$p < .0001$

Table 3

Mean Item Difficulties and Contrast Values Ethnic Group Performance on PPVT-R Forms L and M

	Form L				Form M			
Item	Word	Anglo American p	Mexican American p	Contrast Value	Anglo American p	Mexican American p	Contrast Value	Word
20		1.00	1.00	-.09	1.00	1.00	-.08	
21		1.00	1.00	-.09	1.00	1.00	-.08	
22		1.00	1.00	-.09	1.00	1.00	-.08	
23		1.00	.97	.04	1.00	1.00	-.08	
24		1.00	.97	.04	1.00	1.00	-.08	
25		1.00	1.00	-.09	1.00	1.00	-.08	
26		1.00	.97	.04	1.00	1.00	-.08	
27		1.00	1.00	-.09	1.00	1.00	-.08	
28		1.00	1.00	-.09	1.00	1.00	-.08	
29		1.00	1.00	-.09	1.00	1.00	-.08	
30		1.00	1.00	-.09	1.00	1.00	-.08	
31		1.00	.97	.04	1.00	1.00	-.08	
32		1.00	1.00	-.09	1.00	1.00	-.08	
33		1.00	1.00	-.09	1.00	1.00	-.08	
34		1.00	1.00	-.09	1.00	1.00	-.08	
35		1.00	1.00	-.09	1.00	1.00	-.08	
36		1.00	1.00	-.09	1.00	1.00	-.08	
37		1.00	.07	.04	1.00	1.00	-.08	

Table 3 (continued)

Item	Form L Word	Form L Anglo American p	Form L Mexican American p	Form L Contrast Value	Form M Anglo American p	Form M Mexican American p	Form M Contrast Value	Form M Word
38		1.00	1.00	-.09	1.00	1.00	-.08	
39		1.00	.97	.04	1.00	1.00	-.08	
40		1.00	1.00	-.09	1.00	1.00	-.08	
41		1.00	.97	.04	1.00	1.00	-.08	
42		1.00	1.00	-.09	1.00	.98	.05	
43		1.00	1.00	-.09	1.00	1.00	-.08	
44		1.00	1.00	-.09	1.00	.98	.05	
45		1.00	.97	.04	1.00	.98	.05	
46		1.00	.97	.04	1.00	.97	.05	
47		1.00	.97	.04	1.00	1.00	-.08	
48		1.00	1.00	-.09	1.00	.97	.05	
49		.98	.97	-.06	1.00	1.00	-.08	
50		1.00	.97	.04	1.00	1.00	-.08	
51		1.00	.97	.04	.98	.95	.04	
52		.98	.98	-.09	.97	.90	-.07	
53		.96	.94	-.04	.98	1.00	-.18	
54		.99	.91	.19	1.00	.94	.12	
55		.95	.87	.09	.98	.97	-.07	
56	bark	.97	.82	.32*	.93	.94	-.12	

Table 3 (continued)

Item	Form L Word	Form L Anglo American p	Form L Mexican American p	Form L Contrast Value	Form M Anglo American p	Form M Mexican American p	Form M Contrast Value	Form M Word
57		.94	.92	.01	.98	.97	-.05	
58		.91	.83	.07	.98	.97	-.05	
59	disappointment	.86	.94	-.30*	.98	.97	-.05	
60		.93	.94	-.13	.96	.91	.06	
61		.93	.80	.19	.94	1.00	-.27*	walrus
62	reel	.88	.64	.30*	.97	.97	-.08	
63		.94	.91	.05	.93	.92	-.13	
64		.89	.89	-.07	.95	.91	.00	
65		.87	.86	-.02	.90	.76	.19	
66		.90	.73	.22	.90	.95	-.24	
67		.89	.76	.11	.96	.95	-.06	
68		.91	.87	-.12	.92	.87	.09	
69		.85	.79	-.02	.93	.84	.15	
70	ceremony	.96	.77	.37*	.86	.57	.41*	ceiling
71	casserole	.84	.36	.55*	.86	.87	-.11	
72	vehicle	.91	.74	.33*	.95	.89	-.03	
73		.91	.84	-.01	.86	.79	.16	
74		.84	.75	.13	.80	.54	.16	
75		.71	.68	-.12	.70	.67	-.13	

Table 3 (continued)

Item	Form L Word	Form L Anglo American p	Form L Mexican American p	Form L Contrast Value	Form M Anglo American p	Form M Mexican American p	Form M Contrast Value	Form M Word
76		.72	.49	.22	.84	.71	.19	
77		.80	.78	-.14	.81	.52	.37*	construction
78		.53	.25	.15	.81	.79	-.04	
79		.75	.66	.12	.86	.66	.33*	grain
80		.73	.56	.15	.78	.60	.21	
81		.68	.43	.26	.79	.48	.37*	sorting
82		.69	.63	-.03	.82	.61	.26	
83	demolishing	.68	.39	.29*	.70	.50	.19	
84		.62	.61	-.05	.61	.55	-.01	
85		.63	.32	.28	.66	.51	.11	
86	amazed	.63	.70	-.40*	.53	.37	.11	
87		.45	.37	-.01	.68	.61	.03	
88		.53	.47	.00	.63	.34	.27*	snarling
89		.52	.46	-.01	.60	.44	.11	
90		.45	.31	.07	.43	.26	.06	
91		.45	.36	.02	.44	.25	.12	
92		.46	.29	.07	.41	.36	-.01	
93		.52	.48	-.01	.50	.36	.11	
94		.43	.21	.09	.37	.26	.01	

Table 3 (continued)

	Form L				Form M			
Item	Word	Anglo American p	Mexican American p	Contrast Value	Anglo American p	Mexican American p	Contrast Value	Word
95		.45	.28	.12	.42	.22	.12	
96		.42	.37	-.03	.27	.09	.06	
97		.53	.43	.01	.34	.13	.08	
98		.41	.28	.05	.38	.22	.03	
99		.36	.22	.02	.19	.12	-.06	
100		.33	.20	.07	.32	.08	.10	
101		.26	.14	.02	.21	.07	.02	
102		.20	.24	-.02	.23	.13	-.04	
103		.38	.19	.09	.22	.12	-.03	
104		.34	.18	.07	.26	.12	.01	
105		.28	.16	.02	.17	.05	-.01	
106		.31	.20	.01	.15	.03	-.02	
107		.33	.18	.04	.15	.03	-.03	
108		.28	.10	.11	.09	.00	-.05	
109		.21	.12	-.04	.19	.05	.01	
110		.20	.06	.01	.13	.00	-.02	
111		.09	.08	.08	.12	.03	-.02	
112		.15	.12	.06	.15	.03	.01	
113		.18	.08	-.04	.12	.03	-.03	
114		.18	.11	-.04	.13	.00	-.02	

Table 3 (continued)

Item	Form L Word	Form L Anglo American p	Form L Mexican American p	Form L Contrast Value	Form M Anglo American p	Form M Mexican American p	Form M Contrast Value	Form M Word
115		.25	.09	.01	.08	.00	-.04	
116		.15	.03	-.02	.10	.00	-.03	
117		.09	.03	-.06	.01	.03	-.09	
118		.12	.07	.06	.10	.00	-.04	
119		.16	.09	-.03	.03	.00	-.07	
120		.11	.04	-.06	.08	.00	-.04	
121		.02	.02	-.09	.05	.00	-.06	
122		.07	.00	-.07	.05	.00	-.05	
123		.10	.07	-.07	.05	.00	-.07	
124		.04	.03	-.07	.06	.00	-.06	
125		.05	.00	-.07	.01	.00	-.08	
126		.04	.00	-.08	.03	.00	-.07	
127		.04	.03	-.09	.00	.00	-.08	
128		.08	.00	-.05	.06	.00	-.06	
129		.02	.00	-.09	.02	.00	-.08	
130		.00	.00	-.09	.05	.00	-.07	
131		.15	.00	-.08	.03	.00	-.07	
132		.05	.00	-.08	.00	.00	-.07	
133		.04	.00	-.07	.01	.00	-.07	

Table 3 (continued)

Item	Form L Word	Form L Anglo American p	Form L Mexican American p	Form L Contrast Value	Form M Anglo American p	Form M Mexican American p	Form M Contrast Value	Form M Word
134		.01	.00	-.09	.03	.00	-.07	
135		.00	.00	-.09	.01	.00	-.07	
136		.02	.00	-.09	.02	.00	-.07	
137		.00	.00	-.09	.02	.00	-.07	
138		.02	.00	-.09	.03	.00	-.07	
139		.02	.00	-.09	.02	.00	-.07	
140		.01	.00	-.09	.00	.00	-.07	

Note. Item difficulties were subjected to arcsin transformations before the interaction contrasts were performed.

Significant positive contrasts indicate higher performance by the Anglo group on a particular item, while significant negative contrasts indicate higher Mexican-American performance.

*$p < .05$